Measures of Center and Dispersion

- **Measures of Center**

1. Mean ($\mu, \bar{x}$) - average (equal to the sum of the values divided by the number of values).

a. Population mean, $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$, where N is the population size.

b. Sample mean, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, where n is the sample size.

Example 1: Consider a random sample of 12 monthly salaries (in thousands of dollars):

$$\{2.5, 3.2, 3.2, 3.4, 3.5, 3.6, 3.9, 3.9, 3.9, 4, 4.2, 4.2\}$$

Find the sample mean.

Practice 1: Consider the random sample of 21 length of stay in a hospital (in days):

$$\{1, 1, 2, 3, 4, 5, 5, 6, 6, 7, 9, 9, 9, 10, 12, 13, 13, 15, 18, 24, 28\}$$

Determine the sample mean.

Consider the problem in Example 1. Suppose a 13th person was sampled and the person earns \$50K per month (or \$600K per year). Compute the sample mean. Is the new sample mean a good representative of the sample?

Remark 1: Means are *sensitive* to outliers.

2. Median ($\tilde{\mu}, \tilde{x}$) - middle value (when values are arrangement from lowest to highest). If there are an odd number of values, there will be one value right in the middle. If there are an even number of values, then the median is the average of the middle pair.

Example 2: Using the random sample of 12 monthly salaries, find the sample median.

Practice 2: Using the random sample of 21 lengths of stay, find the sample median.

Consider the problem in Example 2. Again, suppose a 13th person was sampled and the person earns \$50K per month (or \$600K per year). Compute the sample median. What can you say about the magnitude of the effect of the outlying value to the sample median?

Remark 2: Medians are *robust* against outliers.

3. Mode - most frequent occurring value.

Example 3: Using the random sample of 12 monthly salaries, determine the mode.

Practice 3: Using the random sample of 21 lengths of stay, determine the sample mode.

Example 4: Compute the mean, median, and mode for the following data sets:

a. $S_1 = \{1, 2, 5, 5, 8, 9\}$
b. $S_2 = \{3, 4, 5, 5, 6, 7\}$
c. $S_3 = \{5, 5, 5, 5, 5, 5\}$

Remark 3: Locating the center of the data is not enough. We also need to measure the spread of the values.

- **Measures of Dispersion (Spread)**

1. Range = Maximum value - Minimum value
Example 5: Compute the range for the three data sets.

2. Variance

a. Population variance, $\sigma^2 = \dfrac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

b. Sample variance, $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Example 6: Compute the variance for data set 1.

Computing formula: $S^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}(\sum_{i=1}^{n} x_i)^2\right)$

Proof:

Example 7: Recompute the variance for data set 1 using the computing formula.

3. Standard Deviation = $\sqrt{\text{Variance}}$
Example 8: Compute the standard deviation for data set 1.

Homework.

Section 1.3: (pp. 34 - 35) # 33, 37, 39, 41.
Section 1.4: (pp. 43 - 45) # 45, 47, 51, 53(a).